UNIVERSITÉ DE MONTRÉAL

CENTRE DE RECHERCHE SUR LES TRANSPORTS

CASE POSTALE 6128, SUCCURSALE A, MONTRÉAL, QUÉBEC, H3C 3J7

A DISTRIBUTED "LAG" FOR SPATIALLY CORRELATED RESIDUALS

by

Ulrich C.H. Blum*

* Universität Karlsruhe, Institut für Wirtschaftspolitik und Wirtschaftsforschung, P.O. Box 6980, D-7500 Karlsruhe.

The author is indebted to the Alexander von Humboldt Foundation for supporting this research during his stay as Alexander von Humboldt Scholar at the Université de Montréal in 1986/87. He thanks J.-M. Dufour and M. Gaudry for their helpful comments.

Cette étude a été publiée grâce à une subvention du fonds F.C.A.R. pour l'aide et le soutien à la recherche.

Centre de recherche sur les transports - Publication #512
Département de sciences économiques - Cahier #8711
Université de Montréal

Institut für Wirtschaftspolitik und
Wirtschaftsforschung - Discussion Paper 4/87
Universität Karlsruhe

March 1987

Dépôt légal - 1er trimestre 1987

0.1 Abstract

We propose a regression model accounting for spatial correlation among residuals where the data decides on the level of impact of increasing degrees of contiguity.

Key-words: Regression, correlation, space, region, residue impact criteria, naturally ordered data

0.2 Zusammenfassung

Wir stellen ein Regressionsmodell vor, in dem räumliche Autokorrelation der Residuen berücksichtigt wird und die Daten über die Stärke des Einflusses zunehmender Grade der Nachbarschaft entscheiden.

Schlüsselworte: Regression, Korrelation, Raum, Region, Residualeinflußkriterium, natürlich geordnete Daten

0.3 Résumé

Nous proposons un modèle de regression avec autocorrelation spatiale des résidus où les données decident de l'influence de degrés de contiguité croissants.

Mots-clés: Regression, correlation, espace, région, critère d'impact residuel, données en ordre naturel

TABLE OF CONTENTS

	page
0.1 Abstract	I
0.2 Zusammenfassung	I
0.3 Résumé	I
1. INTRODUCTION	1
2. THE BASIC NORIC MODEL	1
3. THE EXTENDED NORIC MODEL	3
3.1 Assumptions	3
3.2 Mathematical formulation	4
3.3 Properties of the distributed spatial autocorrelation formula	5
3.4 A note on the likelihood function	8
4. NUMERICAL EXAMPLE	8
5. CONCLUSION	11
REFERENCES	12
Figure 1: Spatial structure of a regional system	8

1. INTRODUCTION

The increased attention for extracting systematic information from the residuals of a regression equation --both efficient from a numerical and a statistical point of view-- has led to the definition of DORIC- and NORIC-type models (BLUM and GAUDRY, 1986) distinguishing between residue impact criteria based on data directly ordered by the researcher as opposed to naturally ordered data. In its most general form, residue impacts are defined by a matrix W_{nxn} which relates the impact of one or several residues $j = 1,2,\ldots,n$ to residue i. In this paper we will draw attention to a spatial NORIC model formulated in a way that data can decide on the impact of increasing degrees of contiguity in a regional setting by defining the spatial analog of a geometrically distributed "lag" structure known from time series analysis. This seems reasonable both from an economic and a statistical point of view as an alternative to specifying a full impact matrix (matrix of contiguity), R, containing $r_{ij} = 1$, $i \neq j$, and $r_{ii} = 0$, $i = 1,2,\ldots,n$, $j = 1,2,\ldots,n$, and as an alternative to using a pattern of a priori weights.

2. THE BASIC NORIC MODEL

Let $(Y, \underline{X}, w)$ be a (1+m+1) dimensional discrete stochastic process with (1+m+1) tupels of random variables; their realizations are given by $(Y_i, X_{1i}, X_{2i}, \ldots, X_{mi}, v_i)$, $i = 1,2,\ldots,n$, where $Y_i, X_{1i}, X_{2i}, \ldots, X_{mi}$ are regional economic observations and v_i white noise residuals. Let us further assume that this process is analysed by means of a standard regression model to which an error component accounting for spatial autocorrelation is added:

$$\underline{Y} = \beta_o + \beta_1 * \underline{X}_1 + \beta_2 * \underline{X}_2 + \ldots + \beta_m * \underline{X}_m + \underline{u}, \qquad (1)$$

$$\underline{u} = \rho * R * \underline{u} + \underline{v} \qquad (2)$$

Obviously

$$u_i = \rho * \sum_{j=1}^{n} r_{ij} * u_j + v_i \ , \quad i = 1,2,\ldots,n \qquad (3)$$

describes the spatial impact of all observations on observation i by means of the spatial autocorrelation coefficient, ρ, and

$$r_{ij} = \begin{cases} \text{value established by an impact criterion if } i \neq j, \\ 0 \text{ if } i = j \ . \end{cases} \qquad (4)$$

If, for instance, we assume that

$$\left[\begin{array}{l}\text{Each region is correlated with its} \\ \text{contiguous neighbors}\end{array}\right] \qquad \text{(NORIC-1)}$$

and region i=5 is contiguous to regions j, $j \in (1,3,6,9)$ out of a total of n=10 regions, then the fifth line of the normalized matrix R reads as follows:

$$r_5 = (0.25,\ 0,\ 0.25,\ 0,\ 0,\ 0.25,\ 0,\ 0,\ 0.25,\ 0). \qquad (5)$$

As we see here, we may wish to normalize the impact matrix by dividing each line by its sum in order to achieve a convex likelihood function for $\rho \in (-1,1)$ (BOLDUC, 1987); the underlying non-normalized matrix W has the following fifth line:

$$w_5 = (1,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 1,\ 0). \qquad (6)$$

Of course it is clear that higher degrees of contiguity may also be important for explaining the residual $\underline{u}$. However, from a practical point of view, it remains open what function of contiguity, distance or interaction in space or between

regions should be applied to achieve a full matrix R^*, if we state:

[Each residue is correlated with its contiguous neighbors according to a function of interaction] (NORIC-2)

In order to bypass this problem, one might wish to use a formula including different orders of autocorrelation:

$$\underline{u} = \rho_1^* R_1^* \underline{u} + \rho_2^* R_2^* \underline{u} + \ldots + \rho_K^* R_K^* \underline{u} + \underline{v} \qquad (7)$$

assuming $k = 1,2,\ldots,K$ different impact matrices and spatial correlation coefficients. This, unfortunately, replaces the problem mentioned above with the difficulty of properly delineating all necessary regional zones, which is a precondition for identifying the different impact matrices.

3. THE EXTENDED NORIC MODEL

3.1 Assumptions

Let us assume that R^* were the true full residue impact matrix. It is then possible to derive it from a (NORIC-1) type matrix of simple contiguity, R, under the following assumptions:

(1) The autocorrelation between every pair of spatial units is representative of the total.

(2) Feedback effects are possible, for instance in the sense that region 1 influences region 2, region 2 influences region 1 and region 3, region 3 influences regions 1,2 and 4.

(3) The level of spatial interaction is a geometrically decreasing function of distance.

Assumption (1) implies that one true structure of autocorre-

lation exists in the stochastic process. If, for instance, a group of zones were lined up (e.g. if we consider regions on the coast) it would not be possible to have neighboring regions that are positively and simulanously regions one zone apart that are negatively correlated. This possibility is not ruled out by formulation (7). Assumption (2) implies that $r^{*}_{ij} \neq 0$ is possible, contrary to what is said for single order autocorrelation in formula (4).

3.2 Mathematical formulation

We propose the following first order autocorrelation process

$$\underline{u} = \rho * (a_o * R * \underline{u} + a_1 * R^2 * \underline{u} + \dots) + \underline{v}, \qquad (8)$$
$$a_i \in [0,1],\ i = 1,2,\dots$$

where the first term in the parentheses describes first degree spatial autocorrelation according to the NORIC-1 formula. By multiplying the impact matrix with itself once, we obtain all neighbors of neighbors, that is second degree contiguity, etc.. We thus differentiate between <u>orders</u> and <u>degrees</u> of contiguity; clearly the order relates to the number of independent correlation coefficents included, which parallels the time series approach, whereas <u>degrees</u> measure the length of the spatial distance accounted for. Following assumption (2), feedback effects are possible, that is, we include from a specific regional point of view all possible "forward linkages" and "backward linkages".

If we assume a geometric weighing system

$$a_i = a^i * (1-a),\ i = 0,1,2,\dots,\infty, \quad a \in (0,1), \qquad (9)$$

which, in fact, is normalized due to

$$\sum_{i=0}^{\infty} a_i = \sum_{i=0}^{\infty} a^i * (1-a) = (1-a) * \sum_{i=0}^{\infty} a^i = \frac{1-a}{1-a} = 1 , \quad (10)$$

we obtain

$$\underline{u} = \rho * (1-a) * \sum_{i=1}^{\infty} a^{i-1} * R^i * \underline{u} + \underline{v} \quad (11a)$$

$$= \rho * (1-a) * \frac{1}{a} * \sum_{i=1}^{\infty} (a * R)^i * \underline{u} + \underline{v} . \quad (11b)$$

With

$$\sum_{i=0}^{\infty} (a * R)^i = I + a*R + (a*R)^2 + (a*R)^3 + \ldots \quad (12a)$$

$$= [I - a*R]^{-1} = I + a*R*[I + a*R + (a*R)^2 + \ldots] \quad (12b)$$

$$= I + a*R*[I - a*R]^{-1} . \quad (12c)$$

we obtain the <u>distributed spatial autocorrelation formula</u>:

$$\underline{u} = \rho * (\frac{1}{a} - 1) * [(I - a*R)^{-1} - I] * \underline{u} + \underline{v} , \quad (13)$$

$$\underline{u} = \rho * (1-a) * R * [I - a*R]^{-1} * \underline{u} + \underline{v} . \quad (14)$$

3.3 Properties of the distributed spatial autocorrelation formula

(1) If a tends to 0 we obtain

$$\underline{u} = \rho * R * \underline{u} + \underline{v} . \quad (15)$$

which follows immediately from setting a = 0 in formula (14). We see that the smaller the value of a, the larger becomes the impact of near regions.

(2) The larger the value of a, the more total degree of

contiguity of each region is determined by its relative position of contiguity with respect to the other regions. Let us explore the extreme case where a tends to 1. With $r_{ij} \in (0,1)\ \forall\ i,j$, and lines adding up to 1 we obtain a stationary state matrix similarily to what is known from Markov chains (see, for instance COX, MILLER, 1965, pp. 76+). Let us define

$$R' = \begin{bmatrix} r_1' & r_2' & \cdot & \cdot & \cdot & r_n' \\ r_1' & r_2' & \cdot & \cdot & \cdot & r_n' \\ r_1' & r_2' & \cdot & \cdot & \cdot & r_n' \\ r_1' & r_2' & \cdot & \cdot & \cdot & r_n' \\ r_1' & r_2' & \cdot & \cdot & \cdot & r_n' \end{bmatrix}, \qquad (16)$$

as stationary matrix of relative position of contiguity of every zone with

$$r_j' = \sum_{i=1}^{n} w_{ij} \Big/ \underbrace{\sum_{i=1}^{n}\sum_{p=1}^{n} w_{ip}}_{2 \times \text{number of borders}} \qquad (17)$$

where $W = (w_{ip},\ i=1,2,\ldots,n,\ p=1,2,\ldots,n)$ as the original non-normalized matrix of contiguity with (0-1) elements. If R' is the stationary matrix for R, we can prove this by showing that

$$R' * R = R' \qquad (18)$$

To show the opposite ($R * R' = R'$) is trivial and irrelevant as multiplying a column vector of identical values with a row vector where the elements sum up to one clearly does not affect the column vector. Any element in matrix R has the form

$$r_{ij} = \frac{w_{ij}}{\sum_{p=1}^{n} w_{ip}} \quad , \quad i = 1,2,\ldots,n;\ j = 1,2,\ldots,n. \tag{19}$$

Any element in matrix R' has the form

$$r'_j = \frac{\sum_{q=1}^{n} w_{qj}}{\sum_{q=1}^{n}\sum_{p=1}^{n} w_{qp}} \quad , \ i = 1,2,\ldots,n. \tag{20}$$

Multiplying the two matrices (that is row i of matrix R' with column j in matrix R) yields for any element in row j and column i:

$$\hat{r}_{ij} = \frac{1}{\sum_{q=1}^{n}\sum_{p=1}^{n} w_{qp}} * \sum_{r=1}^{n}\sum_{q=1}^{n} w_{qr} * \frac{w_{rj}}{\sum_{p=1}^{n} w_{rp}} \quad , \tag{21a}$$

$$\hat{r}_{ij} = \frac{1}{\sum_{q=1}^{n}\sum_{p=1}^{n} w_{qp}} * \sum_{r=1}^{n} w_{rj} * \underbrace{\frac{\sum_{q=1}^{n} w_{qr}}{\sum_{p=1}^{n} w_{rp}}}_{= 1 \text{ due to the symmetry of matrix W}} = r'_j \ . \tag{21b}$$

Of course we can rewrite equation (8) for $a_i = 1$, $i = 0,1,2,\ldots,$

$$\underline{u} = \rho * \frac{1}{n} * \sum_{i=1}^{n} R^i * \underline{u} + \underline{v} \ ; \tag{22}$$

as R^i converges to R' if i tends to infinity, we obtain

$$\lim_{n\to\infty} \frac{1}{n} * \sum_{i=1}^{n} R^i = \lim_{n\to\infty} \frac{1}{n} * (n*R' + D) = R' + \lim_{n\to\infty} \frac{1}{n} * D = R'. \tag{23}$$

We see that the result converges to R' as the error matrix D from our approximation disappears. We are thus able to give bounds for R^* by

$$r^*_{ij} \in [r_{ij} , r'_j] , \; i = 1,2,\ldots,n, \; j = 1,2,\ldots,n . \qquad (24)$$

3.4 A note on the likelihood function

Computing this system using a maximum likelihood approach necessitates the inclusion of the transformation of variables which takes place due to the assumption of autocorrelation. Thus, we must add the Jacobian to the log-likelihood function which, if concentrated on the coefficients β and the variance σ^2, has the following form under the assumption of normally distributed white noise residuals $\underline{v}$:

$$\text{Ln } L = -\frac{n}{2}*[1+\ln(2*\pi)] - \frac{n}{2}*\ln(\sigma^2) + \left|I-\rho*(1-a)*R*(I-a*R)^{-1}\right| \qquad (25)$$

We may compute the determinant using its eigenvalues

$$\left|I-\rho*(1-a)*R*(I-a*R)^{-1}\right| = \prod_{i=1}^{n}[1 - \rho * (1-a) * \lambda_i] , \qquad (26)$$

where λ_i are the eigenvalues of $R*(I - a*R)^{-1}$.

4. NUMERICAL EXAMPLE

Let us assume the following regional setting

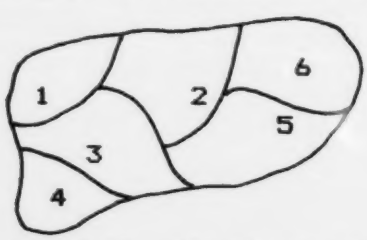

Figure 1: Spatial structure of a regional system

The non-normalized and the normalized matrices of contiguity are

$$W = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 0.00 & 0.50 & 0.50 & 0.00 & 0.00 & 0.00 \\ 0.25 & 0.00 & 0.25 & 0.00 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.00 & 0.25 & 0.25 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.33 & 0.33 & 0.00 & 0.00 & 0.33 \\ 0.00 & 0.50 & 0.00 & 0.00 & 0.50 & 0.00 \end{bmatrix}$$

If we take the matrix R^2 as given below, we see that all elements are full with the exception of r^2_{ij} for (i=4,j=3) and (i=3,j=4) as obviously there is no second degree of contiguity from region 4 to region 3 and, similarily, region 4 can not be reached from region 3 in exactly two steps.

0.2500000	0.1250000	0.1250000	0.1250000	0.2500000	0.1250000
0.0625000	0.3958333	0.2083333	0.0625000	0.1875000	0.0833333
0.0625000	0.2083333	0.5208333	0.0000000	0.0625000	0.1458333
0.2500000	0.2500000	0.0000000	0.2500000	0.2500000	0.0000000
0.1666667	0.2500000	0.0833333	0.0833333	0.3333333	0.0833333
0.1250000	0.1666667	0.2916667	0.0000000	0.1250000	0.2916667

If we take the relative position of each region with respect to other regions, we will find 8 borders or 16 border relations (forth and back). The absolute number of borders of each region is given by

(2, 4, 4, 1, 3, 2)

so that their relative position is

(0.1250, 0.2500, 0.2500, 0.0625, 0.1875, 0.1250).

If matrix R is taken to its eleventh power, it already yields

0.1226865	0.2498177	0.2566432	0.0601623	0.1840474	0.1266429
0.1249089	0.2497185	0.2506308	0.0622701	0.1871739	0.1252979
0.1283216	0.2506308	0.2399874	0.0660515	0.1926994	0.1223093
0.1203246	0.2490803	0.2642060	0.0575158	0.1801176	0.1287557
0.1226983	0.2495652	0.2569326	0.0600392	0.1838978	0.1268669
0.1266429	0.2505958	0.2446185	0.0643779	0.1903004	0.1234646

If we vary the value of a in formula

$$(1 - a) * R * (I - a*R)^{-1},$$

we can clearly follow the convergence process already established in part 3.3 of the paper. A small value for a approximately yields the normalized matrix of contiguity containing only first degree autocorrelation:

For a = 0.001:

0.0002498	0.4996251	0.4996252	0.0001249	0.0002499	0.0001250
0.2498126	0.0003956	0.2499584	0.0000625	0.2499375	0.2498334
0.2498126	0.2499584	0.0005204	0.2497501	0.2498127	0.0001458
0.0002498	0.0002500	0.9990005	0.0002498	0.0002498	0.0000001
0.0001666	0.3332500	0.3330836	0.0000833	0.0003331	0.3330834
0.0001250	0.4996668	0.0002915	0.0000001	0.4996251	0.0002915

For a = 0.01:

0.0024814	0.4962676	0.4962738	0.0012407	0.0024876	0.0012490
0.2481338	0.0039376	0.2495880	0.0006240	0.2493755	0.2483411
0.2481369	0.2495880	0.0051669	0.2475129	0.2481441	0.0014511
0.0024814	0.0024959	0.9900516	0.0024751	0.0024814	0.0000145
0.0016584	0.3325007	0.3308588	0.0008271	0.0033126	0.3308423
0.0012490	0.4966822	0.0029022	0.0000073	0.4962634	0.0028959

Increasing the value of a lets the the matrix converge to a stationary state:

For a = 0.5:

0.0879645	0.3487521	0.3549639	0.0443705	0.1031614	0.0607876
0.1743760	0.1554077	0.2396007	0.0299501	0.2196339	0.1810316
0.1774820	0.2396007	0.1802551	0.1475319	0.1930116	0.0621187
0.0887410	0.1198003	0.5901276	0.0737659	0.0965058	0.0310593
0.0687743	0.2928453	0.2573489	0.0321686	0.1247920	0.2240710
0.0607876	0.3620632	0.1242374	0.0155297	0.3361065	0.1012757

For a = 0.95:

0.1230488	0.2579803	0.2601201	0.0617785	0.1790904	0.1179823
0.1289901	0.2426713	0.2478135	0.0588557	0.1910395	0.1306303
0.1300600	0.2478135	0.2471761	0.0712043	0.1859930	0.1177535
0.1235570	0.2354228	0.2848173	0.0676441	0.1766933	0.1118658
0.1193936	0.2547194	0.2479906	0.0588978	0.1836734	0.1353258
0.1179823	0.2612606	0.2355070	0.0559329	0.2029886	0.1263291

A value of a = 0.95 nearly results in the matrix

$$\lim_{a \to 1} (1-a) * R * (I - a*R)^{-1}$$

where all columns contain identical values.

5. CONCLUSION

We have established a system to account for first order spatial autocorrelation where data decides on the strength of an interaction parameter which is the spatial analog of the weighing parameter defined in classical time-series polinomial lag structures. It permits the easy inclusion of all spatial relations based on a simple matrix of first order contiguity.

REFERENCES:

BLUM, U. and M. GAUDRY, 1986, An Example of Correlation among Residuals in Directly Ordered Data, University of Karlsruhe, Institut für Wirtschaftspolitik und Wirtschaftsordnung Discussion Paper 3/86, Université de Montréal, Département de sciences économiques # 8619, Centre de recherche sur les transports # 456.

BOLDUC, D., 1987, On the Estimation of Models with Generalized SAR (1) Processes on the Residuals of a Regression, Université de Montréal, Centre de recherche sur les transports # 508.

COX, D.R. and H.D. MILLER, 1965, The Theory of Stochastic Processes, Methuen & Co. Ltd., London.